Catalog Augmentation for Data Mining and User Presentation of Mined Data

Contributing Team Members:

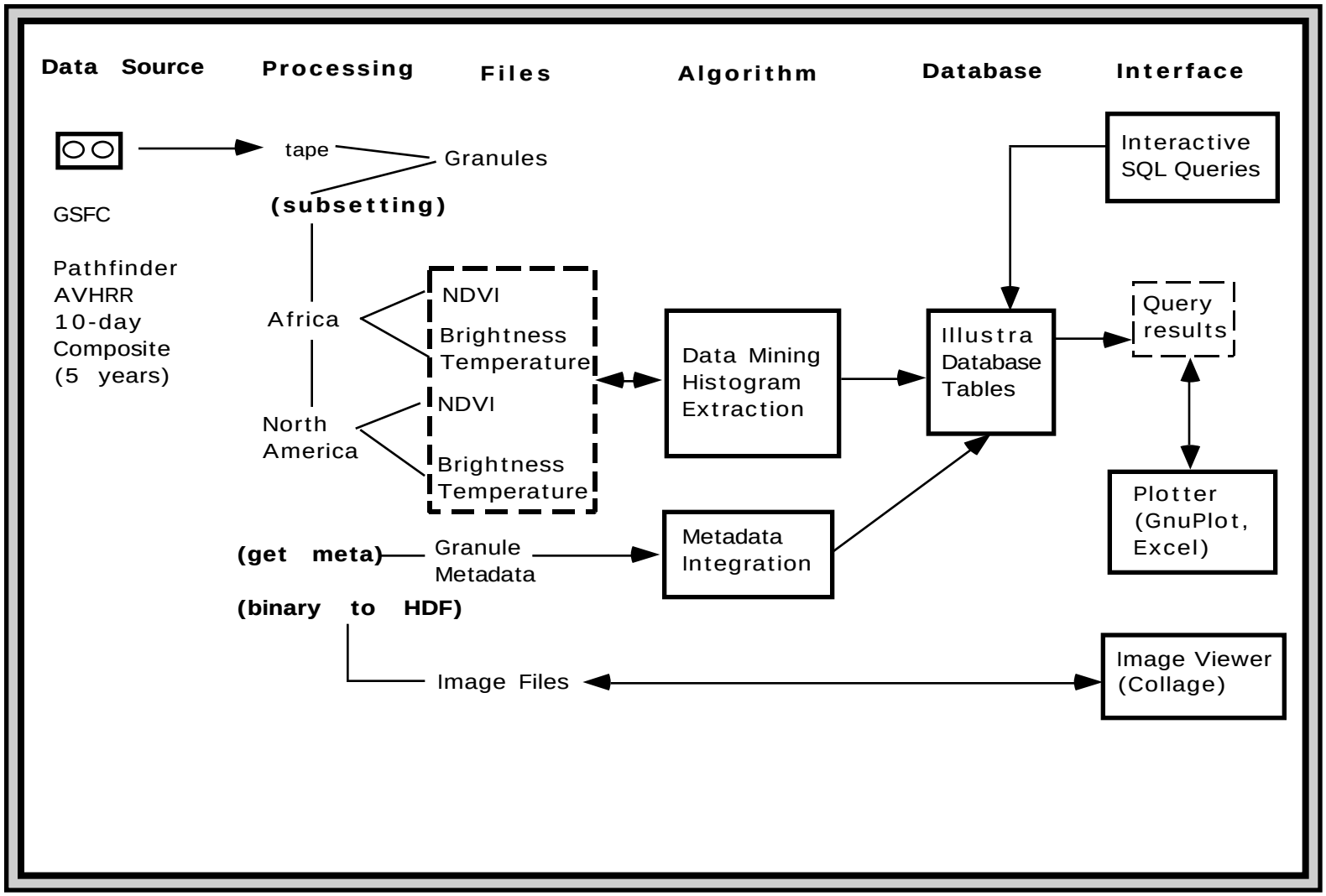
Mike Moore	NASA GSFC
Robin Pfister	NASA GSFC
Robert Harberts	HSTX
Patrick Agbu	HSTX
Ruiming Chen	HSTX
Jay Jong	HSTX
Mike Goldberg	MITRE

Data Mining Investigations

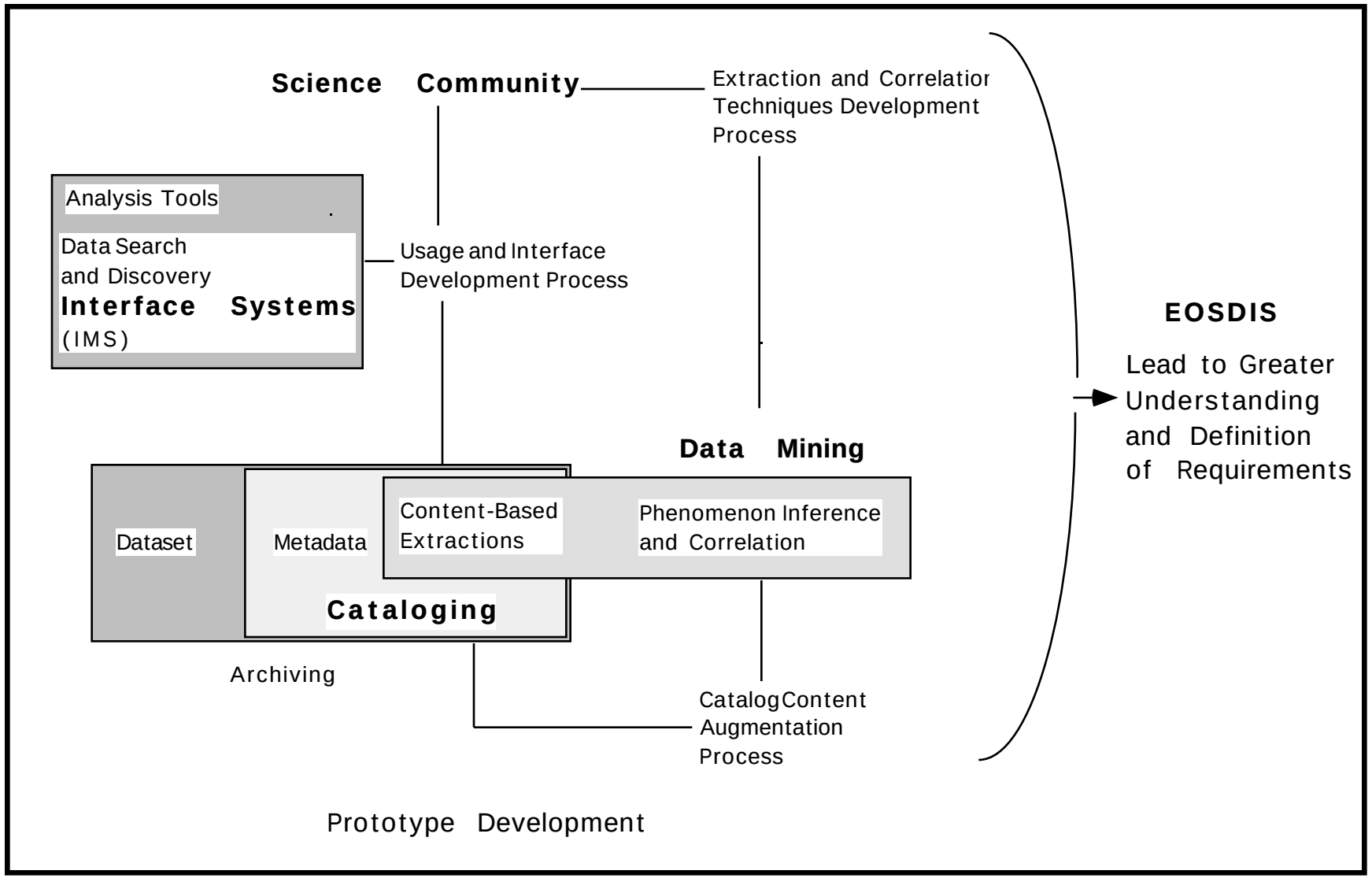
- Study how to support iterative development of content-based data search and analysis techniques by scientists.

- Explore how these techniques can be delivered to the science community
  - Achieve greater understanding of the interrelationships between content-based metadata and user interface tools
  - Learn how others will be able to develop and add their own analysis routines and metadata

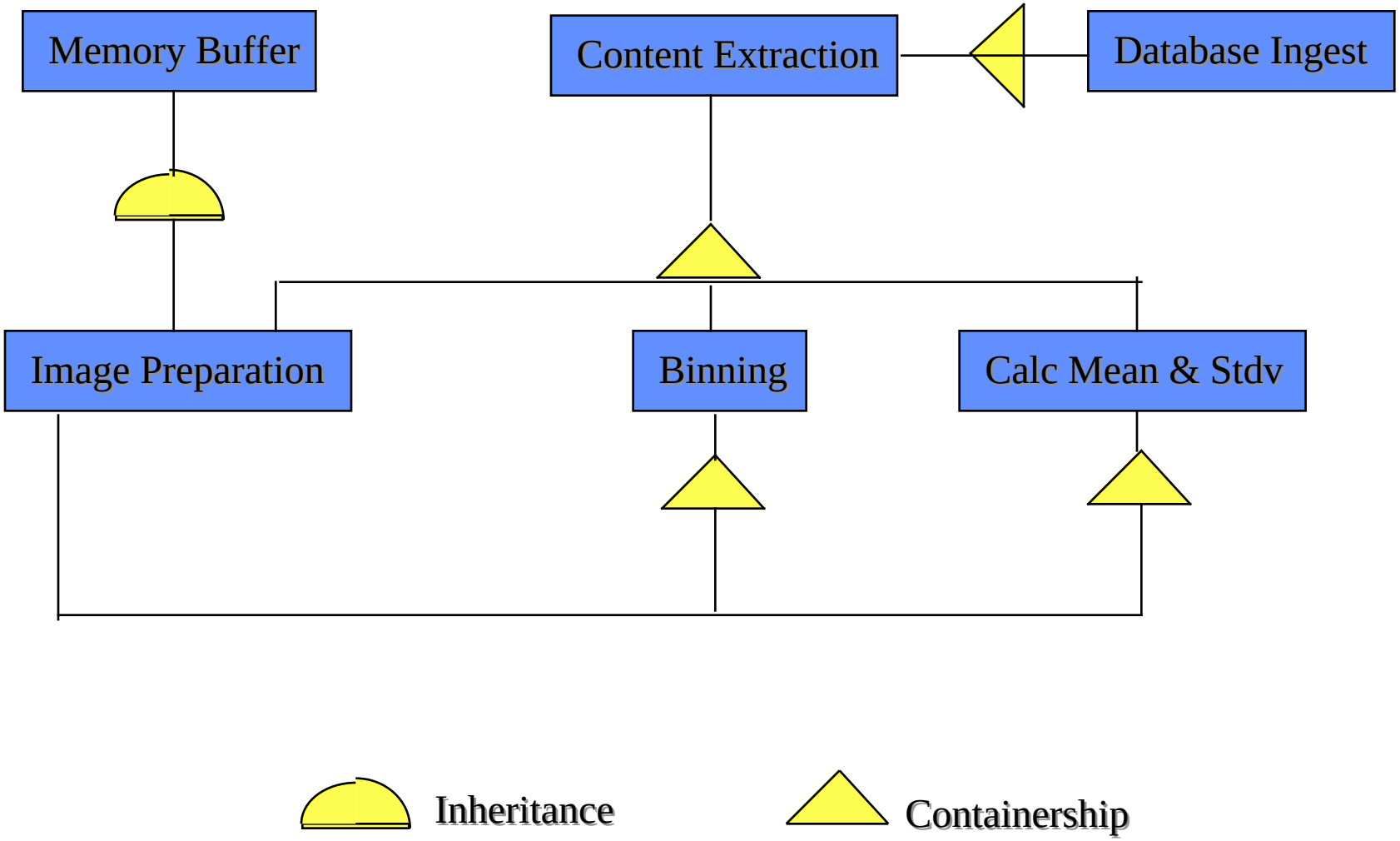
Prototype System Established to Support Image Content Extraction and Science Study



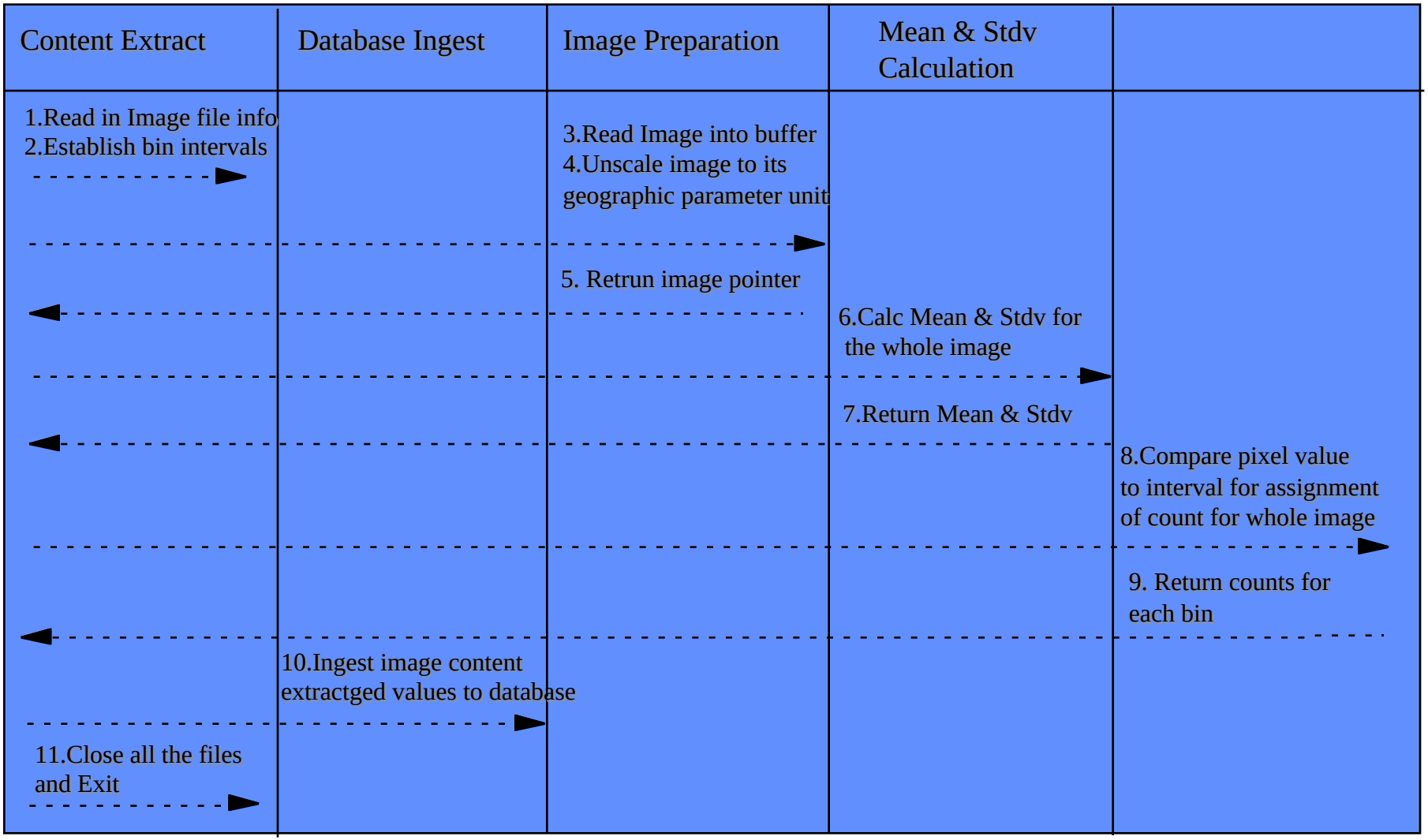
Data Mining Project Context



Static View of the Content Based Extraction System



Dynamic View of the Content Based Extraction Sy stem



Mining Anomalous Geophysical Characteristics from Large Image Data Collections

Objectives

- Augment image database catalog with content-based metadata
- Use cataloged metadata to locate satellite images containing characteristics of severe soil moisture deficit condition (drought)
- verify that mined geophysical parameters are linked to images related to drought.

Approach

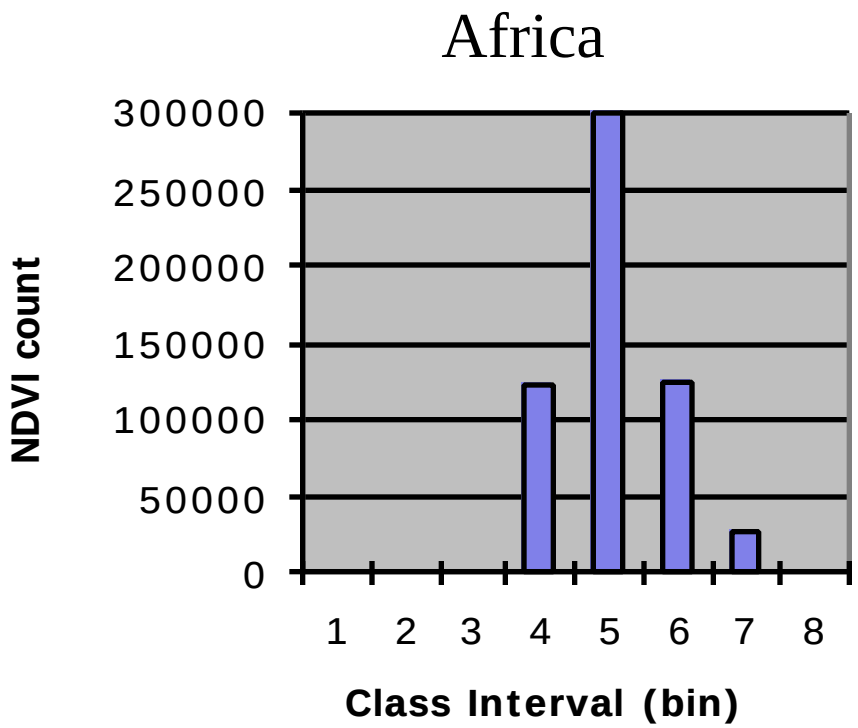
Our data mining approach has the following features:

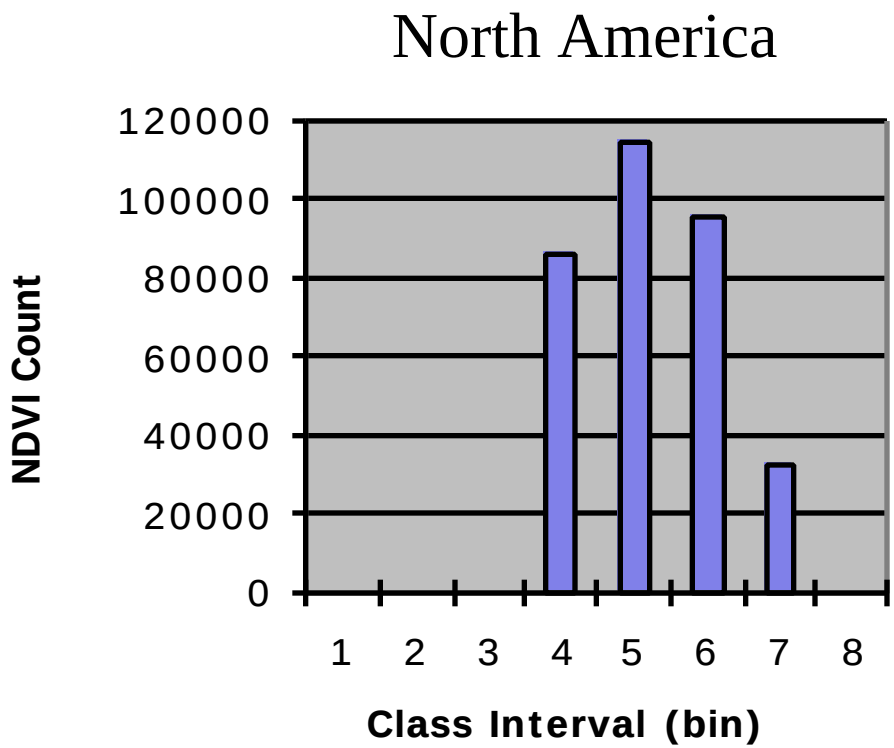
- Information extraction before initiation of search – to ensure a system that was practical and cost-effective for large image data sets
- Use of a simple generic information measure – to emphasize the system concept rather than technique development
- Iterative query by exact match or user defined thresholds – to take advantage of the scientist's knowledge of the parameter characteristics and interaction with the biosphere.

Materials and Methods

- Source Data
- Subsetting
- Algorithm for binning pixel counts
- Database catalog
- Interface tools
- Mining technique and analysis
- Validation

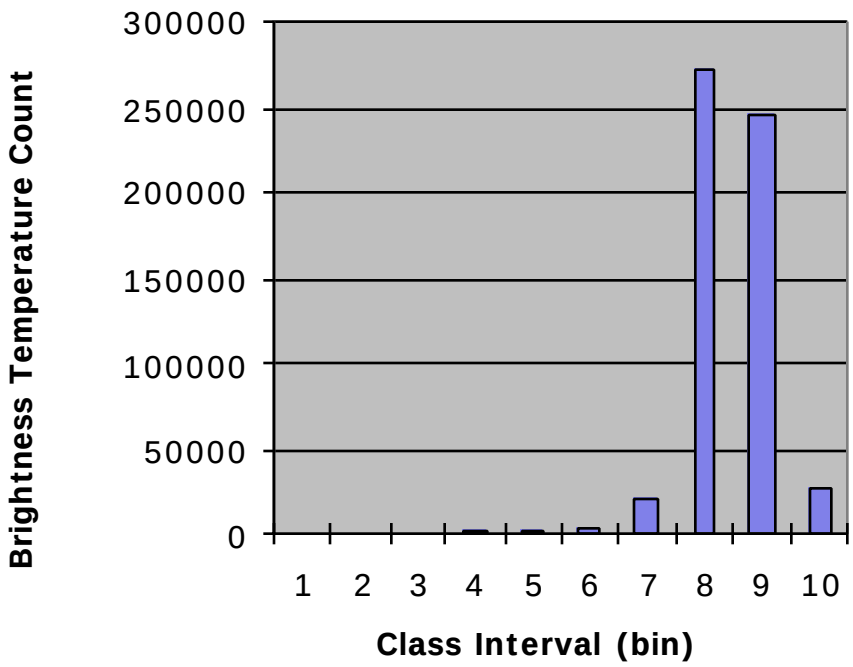
Cumulative average NDVI



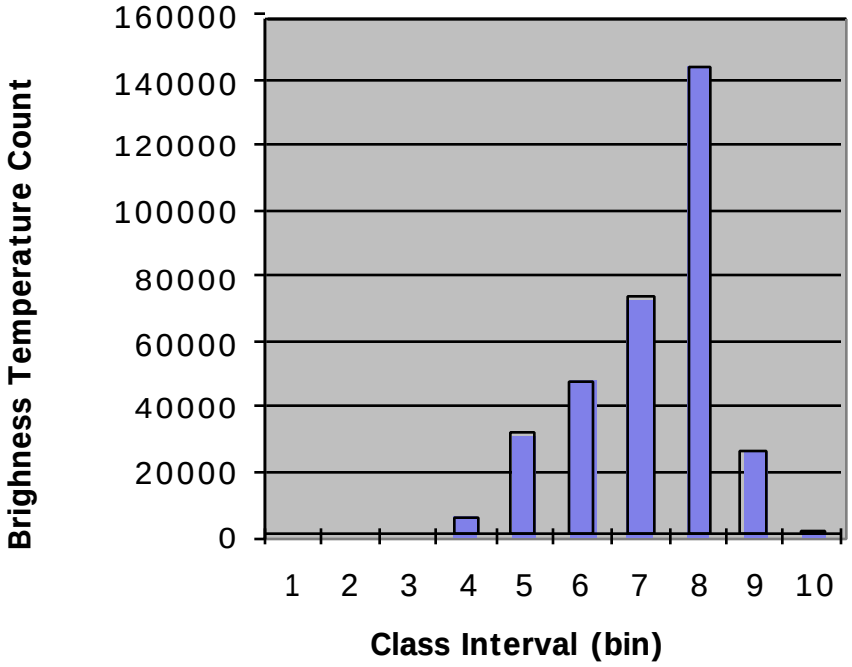


Cumulative average Brightness Temperature

Africa

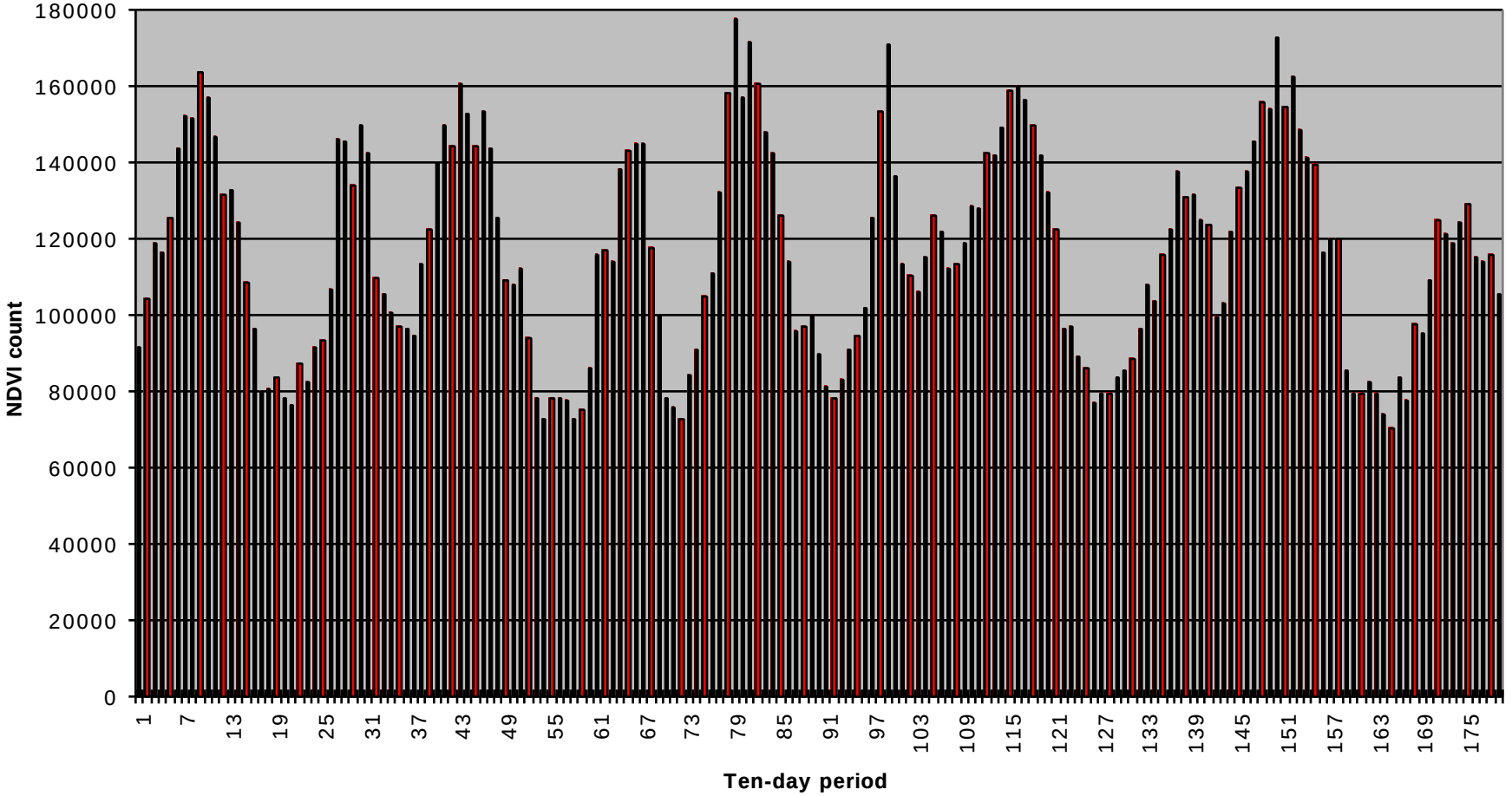


North America

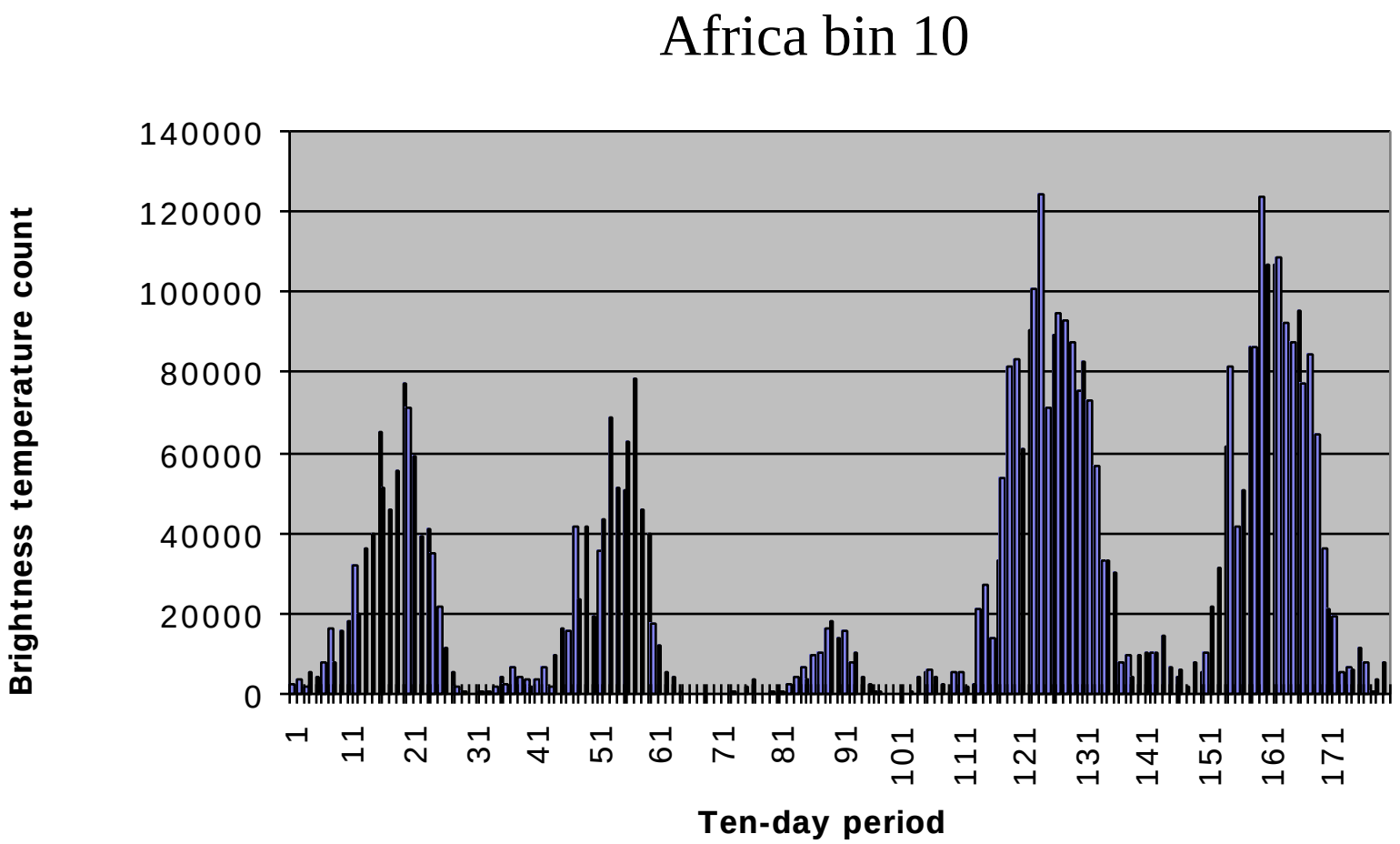


NDVI Timeseries (1986-1990)

North America bin 4



Brightness Temperature Timeseries (1986-1990)



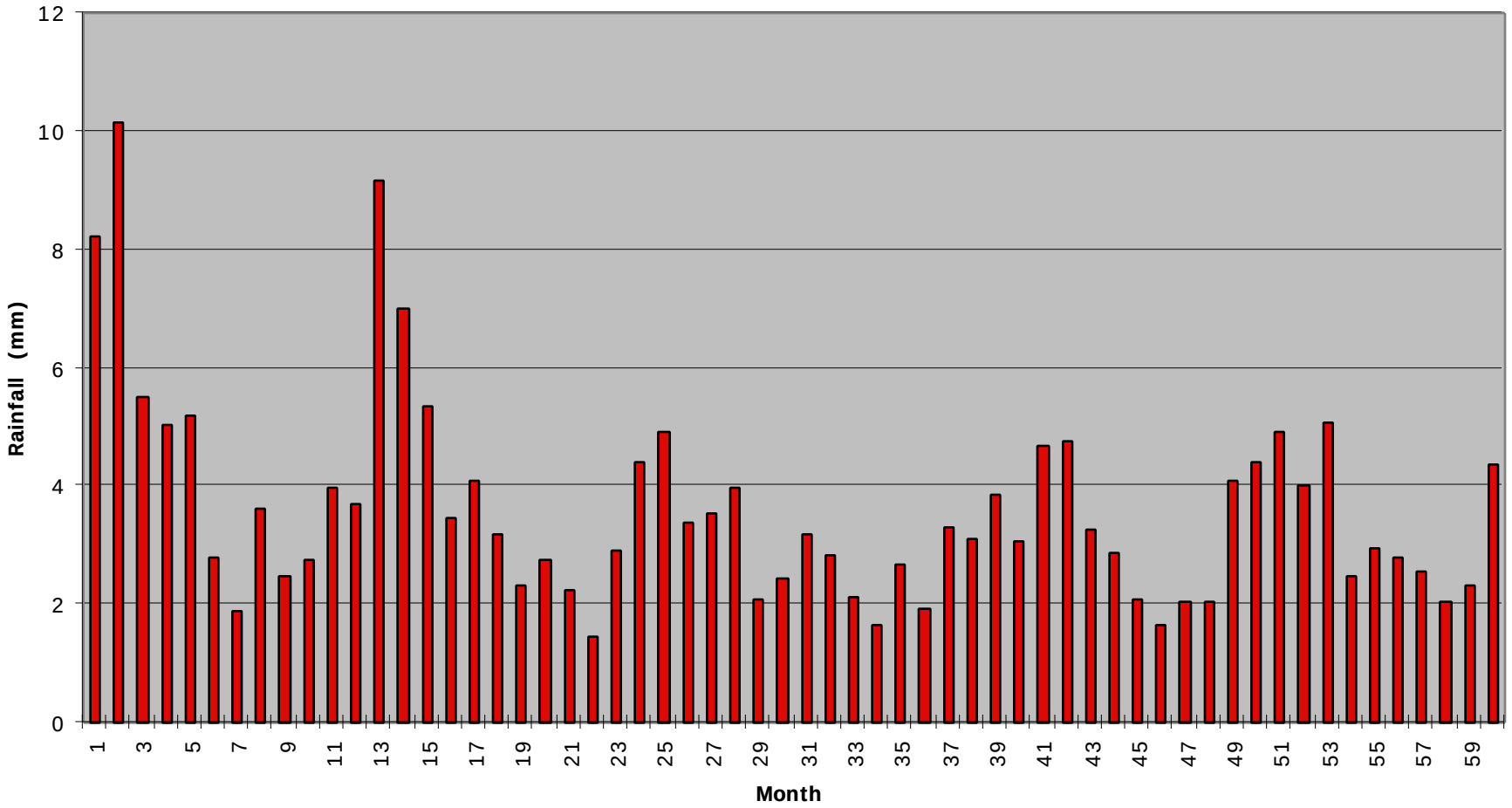
Results

Correlation Coefficients Between Bin Counts & Derived Monthly Rainfall from 1986-1990

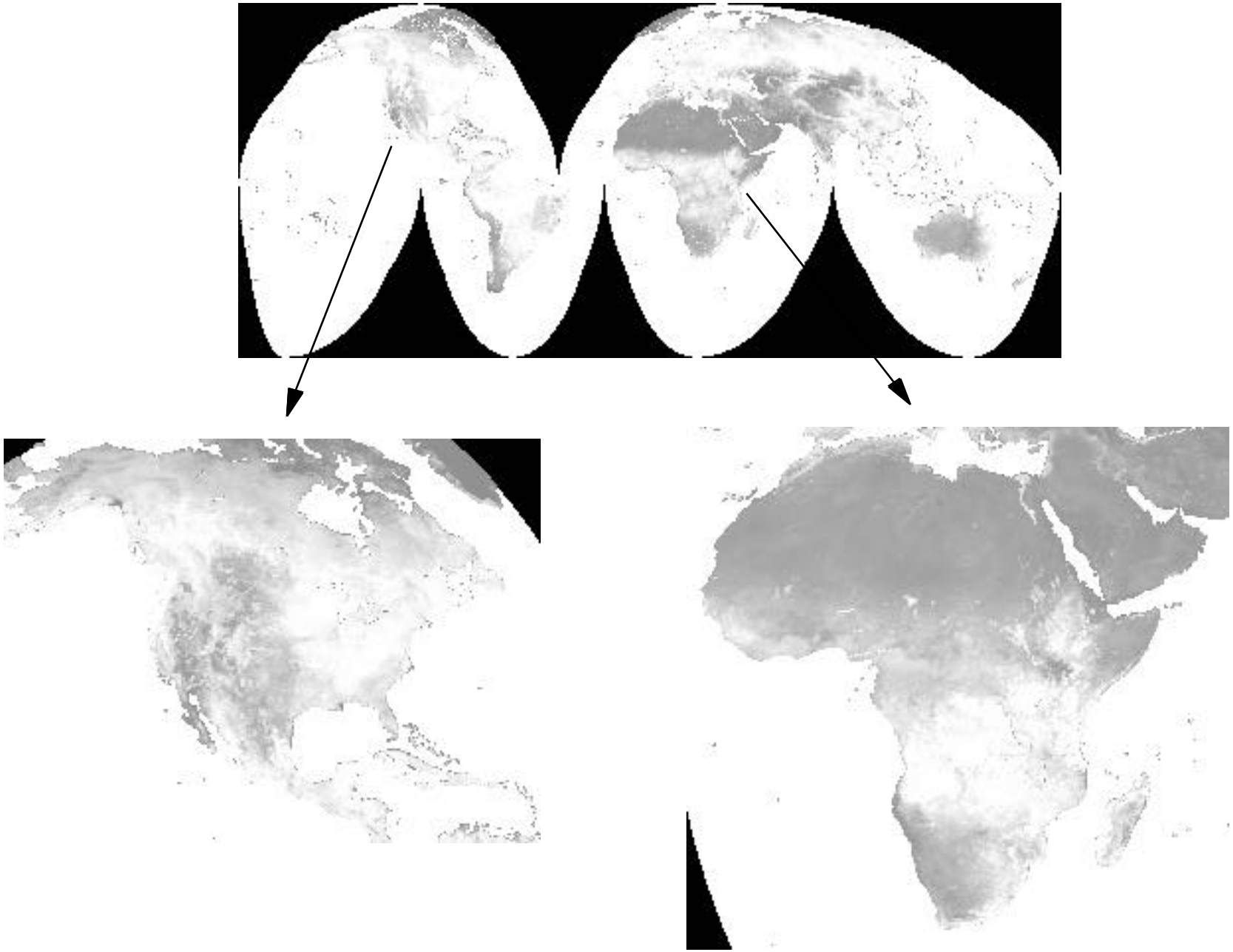
Geophysical parameter (counts)	Derived rainfall (mm)
Africa bin 4 NDVI	0.064
Africa bin 5 NDVI	0.007
Africa bin 8 brightness temperature	0.193
Africa bin 10 brightness temperature	-0.444**
North America bin 4 NDVI	0.496**
North America bin 5 NDVI	0.185
North America bin 7 brightness temperature	-0.148
North America bin 8 brightness temperature	-0.377**

** Statistically significant at the .01 probability level.

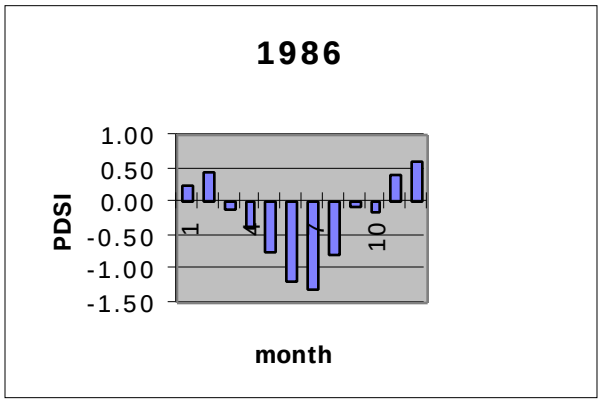
Derived Monthly Rainfall for North America

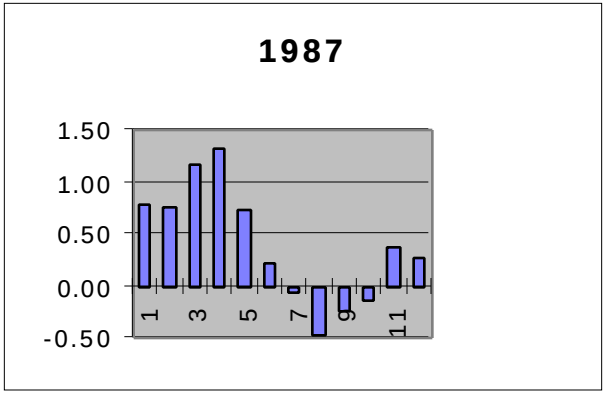


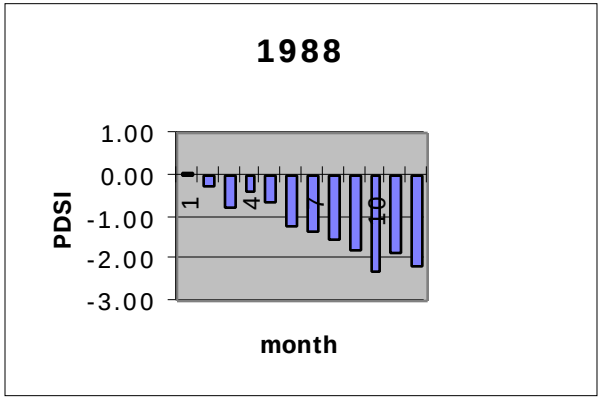
Pathfinder AVHRR Land Composite Data Set

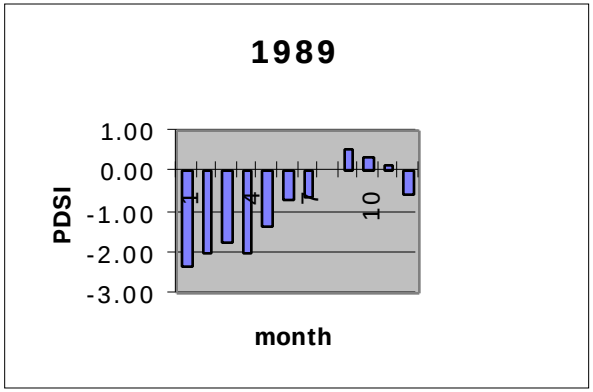


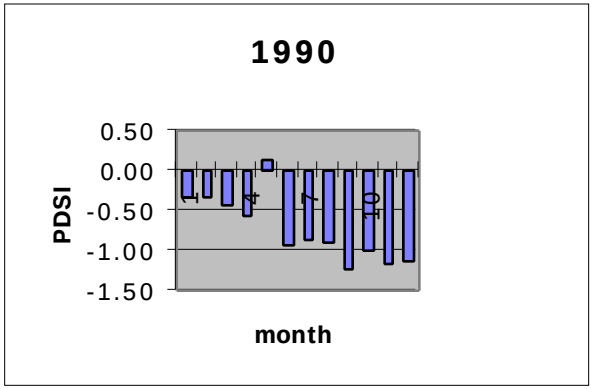
Palmer Drought Severity Indices for Transect Across Continental United States











Summary

- Periods of significant moisture deficit conditions on the land surface were gleaned from analysis of cataloged and mined parameters over time.

- Independent data sets, corroborated the assertion that abnormal rates of change in geophysical processes recorded in imagery, could be identified relatively quickly in large satellite databases.

- Results suggest that other event related data or phenomena could be mined from large data collections
  - — locust infestation (vegetation index)
  - — earthquake (seismic data)
  - — severe storm (atmospheric pressure)
  - — forest fire (vegetation index)
  - — volcanic eruption (atmospheric sulfur dioxide concentration), among others.